

WhatsApp intel·ligent: Anàlisi d'una conversa de WhatsApp

Isaac Serrano Serra

Resum— WhatsApp ha esdevingut una de les principals eines de comunicació informal en els darrers anys. Aquest fet ha comportat un canvi en la qualitat de les nostres comunicacions: ara podem intercanviar informació des del nostre dispositiu mòbil instantàniament i tenim la possibilitat d'accedir-hi en qualsevol moment. Encara que totes aquestes característiques ens faciliten molts aspectes de la vida, també ens obliguen a estar disponibles les 24 hores del dia pendents dels fils de les nostres converses. Aquest projecte s'ha desenvolupat per tal d'ajudar a analitzar i resumir converses en grup en les quals s'hagi produït un gran nombre de missatges on nosaltres no hàgim estat presents a través d'una eina senzilla i fàcil d'utilitzar. A causa de la gran quantitat d'informació que es pot intercanviar en aquesta plataforma i dels diferents tipus (text, icones, imatges, vídeo, veu), aquest projecte està enfocat en converses de grup les quals tinguin objectius molt concrets, com ara decidir llocs, regals, viatges, festes...

Paraules clau—WhatsApp, resum, anàlisi, corrector

Abstract— WhatsApp has become one of the main communication tools in the last years. This fact has led us to a change in the quality of our communications, now we can exchange information instantly from our mobile devices and we have the ability to access to them at any time. Although all these features can facilitate many aspects of our lives, it makes us to be available 24 hours a day, attentive to the threads of our conversations. This project has been developed to help to analyse and summarize group conversations in which there has been a big number of messages and where we were not present through a simple and easy to use tool. Due to the large amount of information that can be exchanged for this platform and the different types (text, icons, images, video, voice), this project is focused on group conversations that have very specific goals, such as deciding sites, gifts, trips, parties...

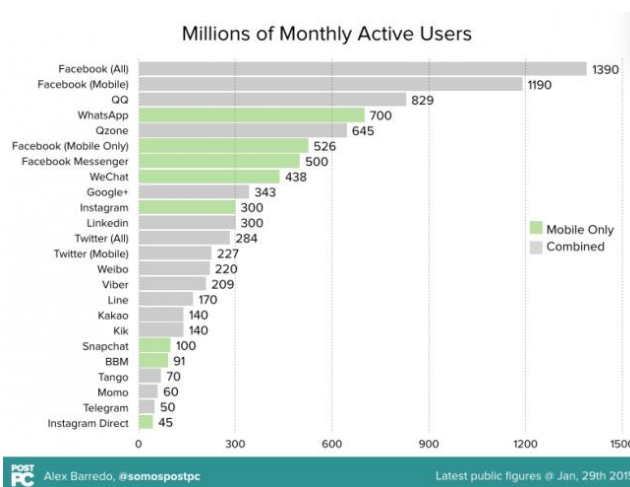
Index Terms— WhatsApp, summary, analysis, corrector

1 INTRODUCCIÓ

EL fenomen de WhatsApp va començar l'any 2009, des de llavors la xifra d'usuaris actius no ha parat de créixer. Al febrer de 2014 l'empresa WhatsApp inc. va ser adquirida per Facebook, un dels altres gegants d'Internet, per la xifra de 19.000 milions de dòlars.

En el mercat també trobem altres opcions que ens ofereixen característiques semblants. Entre aquestes hi podem trobar Telegram, Line o Snapchat que tenen un funcionament molt semblant, o bé altres com Facebook que són xarxes socials però tenen un servei de missatgeria.

La seva facilitat d'ús, la compatibilitat amb diversos sistemes i el seu baix cost¹ ha fet que al gener del 2015



Il·lustració 1 Nombre d'usuaris actius mensuals per a les diferents eines i aplicacions.

hagi assolit 700 milions d'usuaris actius. Com podem contrastar a la il·lustració 1, WhatsApp és l'aplicació que té més usuaris actius mensuals en comunicació mòbil

- E-mail de contacte: Isaac.serranos@e-campus.uab.cat
- Menció realitzada: Tecnologies de la Informació.
- Treball tutoritzat per: Joan Serra Sagristà (Departament d'Enginyeria de la Informació i de les Comunicacions)
- Curs 2014/15

¹ WhatsApp es pot descarregar gratuïtament en algunes plataformes i pagar una petita quantitat per utilitzar el servei.

seguida per Facebook i WeChat. Amb aquestes dades, és difícil trobar usuaris de mòbils intel·ligents que no utilitzin aquest servei. Ens trobem, doncs, en una època en què les comunicacions tendeixen cap a aquests tipus de plataformes i aquest projecte està realitzat per la plataforma més utilitzada però la idea podria estendre's per a altres tipus d'aplicacions en les quals el tipus de conversa que hi trobem és el mateix.

Actualment, la comunicació diària entre grups d'amics se sol basar en intercanviar missatges de text, veu, imatges... Aquest treball se centra en les converses en grup, una altra característica important de l'aplicació. Un grup és una conversa on participen més de dues persones i on els missatges arriben a tothom que hi és inclòs.

És comuna la creació d'aquestes converses entre grups d'amics per mantenir el contacte cada dia. Un dels altres motius pel qual es creen converses és per planejar i discutir sobre un tema o esdeveniment concret. Aquests esdeveniments poden ser, entre d'altres, aniversaris, partits, sopars, regals...

Aquestes últimes converses solen presentar unes característiques comunes:

- Neixen en una data pròxima a l'esdeveniment.
- Tenen un temps de vida d'uns quants dies o setmanes.
- Només es tracta sobre un tema molt concret (tot i que poden sorgir converses paral·leles).
- La majoria dels missatges es produeixen en un període curt de temps.

Derivat d'aquestes característiques, es pot deduir que també poden comportar alguns problemes als usuaris. Suposem que som usuaris de WhatsApp i per alguna raó, hem estat desconnectats tot el dia. Finalment quan entrem en el servei, observem que hem sigut convidats en una conversa multitudinària i en la qual ja hi ha centenars de missatges.

Aquest projecte està enfocat a aquestes situacions. Una eina capaç de proporcionar un resum de la conversa, on es puguin veure les intervencions més importants i unes estadístiques que ens proporcionin informació sobre els usuaris i la conversa, com per exemple, qui ha parlat més o el nombre total de missatges.

Una altra característica que trobem en una conversa de WhatsApp, és la presència de molts errors ortogràfics. No només és exclusiu d'aquest servei sinó que és comú en moltes aplicacions per a dispositius mòbils, ja que l'escriptura en aquests teclats sol ser difícil per culpa de la mida de les pantalles, la proximitat entre tecles, etc. Aquest fet dificulta la tasca per a una màquina, ja que paraules escrites completament diferents, poden tenir el mateix significat.

Les imatges i icones també es troben dins de les con-

verses de WhatsApp. Aquests elements aporten molta informació, per exemple una icona amb una persona somrient pot significar que sí o que estem d'acord amb una proposta. Fins i tot, podem obviar part del text per una imatge, en una situació en què estem escollint un objecte de regal, un usuari pot enviar una fotografia d'un element, si aquest és finalment l'escollit i en cap moment es parla d'aquest element serà impossible d'incloure'l en el resum.

A continuació s'expliquen els objectius del projecte. Des dels objectius proposants en les primeres etapes del projecte fins als objectius finals, explicant tots els canvis i decisions que es van prendre.

1.1 Objectius

En les primeres etapes del projecte, els eren una mica diferents dels actuals. Al principi estava enfocat per desenvolupar una aplicació per a dispositius mòbils amb sistema operatiu Android. Aquest objectiu es va estudiar perquè l'aplicació de WhatsApp s'utilitza sobretot des del mòbil i per a l'usuari sembla més còmode utilitzar els dos serveis des del mateix dispositiu.

A mesura que es va anar estudiant el funcionament de WhatsApp es va descobrir que aquest és de codi propietari, per tant no es pot accedir al seu codi ni tampoc saber com funciona realment. Això dificulta aconseguir una màxima integració amb aquest treball. Durant el procés de recerca d'informació d'altres projectes també es va veure que hi havia més documentació en llenguatges de programació que no eren els utilitzats en el desenvolupament per aplicacions d'Android. Aquests fets van fer replantejar un dels objectius més importants del projecte i finalment es va prendre la decisió de no desenvolupar una aplicació per a mòbils.

Aquest canvi va suposar que ara l'usuari ha de proporcionar manualment les converses (es poden exportar converses en format text des de la mateixa aplicació WhatsApp) per tal de poder analitzar-les. Encara que aquest canvi pugui semblar molt arriscat, en les últimes etapes del projecte va aparèixer WhatsApp Web[1], un client per a navegador que ens permet utilitzar el servei de missatgeria des de l'ordinador.

L'objectiu principal es va veure modificat, ja que era l'aspecte del producte final. Els altres objectius s'han mantingut al llarg del projecte i finalment es poden agrupar en els següents:

- Proporcionar un resum amb les intervencions més importants d'una conversa.
- Corregir les faltes d'ortografia per tal de millorar la qualitat del resum.
- Proporcionar estadístiques en forma de gràfics i taules per obtenir informació sobre la conversa i els seus usuaris.
- Realitzar una interfície que faciliti l'ús de l'aplicació i la seva personalització.

A més a més d'aquests objectius també hi ha alguns requeriments que ha de complir el producte final:

- **Temps d'execució:** ha de ser inferior al temps estimat que trigaria un usuari a llegir tota la conversa.
- **Usabilitat:** L'usuari ha de ser capaç d'utilitzar i entendre que fa el sistema sense ajuda de documentació.

En el següent apartat es mostren alguns dels projectes que s'han estudiat i que han estat el punt de partida d'aquest projecte.

1.2 Estat de l'art

Actualment trobem en el mercat d'aplicacions algunes solucions que ens permeten obtenir estadístiques de la nostra aplicació de WhatsApp.

Per exemple en el Google Play d'Android² trobem aplicacions com WhatStat[2] que ens fa un seguiment dels nostres contactes i converses. Amb aquesta aplicació es pot saber quins són els contactes que parlen més, qui comença més vegades una conversa, a quina hora envia més missatges, etc.

Referent al resum, no s'ha localitzat cap eina específica ni cap treball relacionat amb aquest objectiu. No obstant això, hi ha molta documentació relacionada amb resums de converses entre correus electrònics. Entre aquests dos tipus de converses, existeixen moltes similituds i fan que molts dels mecanismes emprats per als correus puguin ser utilitzats per una conversa en grup de WhatsApp.

Entre aquests projectes s'hi pot trobar l'article de Zajic et al. [3], que presenta moltes similituds amb aquest projecte. En aquest article s'expliquen diferents tècniques per a resumir converses en correus electrònics. El mètode de CMS (Collective Message Summarization) s'aproxima bastant a la nostra casuística, ja que tracten tots els correus com a un únic document i utilitzen eines estadístiques per seleccionar els millors candidats per aparèixer al resum.

Seguint en aquesta mateixa línia, trobem un altre article que parla de resumir converses escrites i orals. Gabriel Murray et al. [4] realitzen diferents estudis sobre converses realitzades en reunions i sobre correus electrònics. En aquest cas també s'utilitzen fórmules estadístiques per tal de puntuar les frases. Entre aquestes mesures hi trobem la "F-score", "precision" o "recall" obtingudes a partir del nombre de participants, nombre d'intervencions, nombre d'intervencions per persona, etc. En aquest article intenten aportar més precisió a l'hora de resumir aquests tipus de textos aportant més característiques a tenir en compte a l'hora d'avaluar una conversa.

A Guanting Tang et al. [5] trobem eines de mineria de dades relacionades amb els correus electrònics per tal de realitzar tasques com filtrar per categoria, detectar spam o identificar persones a partir de les característiques d'un text. Per assolir aquestes tasques utilitza classificadors com ara el tf-idf [6] i el Naïve Bayes, que classifica les paraules segons la probabilitat de pertànyer a una categoria basant-se en unes dades d'aprenentatge anteriors.

Giuseppe Carenini et al. [7] proposa una solució per a la lectura de correus electrònics en dispositius que compten amb la pantalla petita en la qual és difícil la seva lectura. Per a fer-ho identifica en el text les paraules clau que més apareixen en una conversa (tf-idf). Un cop identifica aquestes paraules proposa diferents algorismes per a realitzar resums, com per exemple comparant-les amb el centroid (vector que conté la mitjana d'aparició de totes les paraules del document).

Finalment Owen Rambow et al. [8] exposen un conjunt de característiques a tenir en compte per tal de realitzar un resum en una conversa de correu electrònic. Entre algunes d'aquestes característiques hi trobem la posició de la frase dins del document, si la frase es tracta d'una pregunta, el nombre de respostes que ha obtingut una pregunta o la mitjana en què una paraula apareix en el text. Totes aquestes característiques són importants a l'hora de fer un resum i es poden utilitzar en aquest projecte.

1.3 Metodologia

La metodologia emprada per a desenvolupar aquest projecte ha estat basar-se en projectes similars per tal de tenir un punt de partida. Un cop estudiats com funcionen aquests altres projectes, es van identificar les parts necessàries amb què comptaria la part d'implementació. Per exemple es va trobar una implementació que resumia un paràgraf de text a partir de les paraules més repetides [9]. A partir d'aquí s'han desenvolupat les diferents parts de la implementació i se n'ha controlat la qualitat realitzant diferents controls, tant interns com externs (altres usuaris). Un exemple el podem trobar en el mòdul "resumidor" de la implementació. A mesura que es van començar a obtenir resultats, es va consultar a diferents usuaris que mostressin la seva opinió sobre el resultat.

Finalment es van integrar els diferents mòduls per avaluar com influïen amb els altres i s'han avaluat els resultats realitzant més proves amb diferents usuaris.

En l'apartat 2 es parla de l'estructura que té la part implementada d'aquest projecte i on s'expliquen les diferents parts per poder entendre com funciona i més endavant servirà per comprendre els resultats de l'apartat d'experimentació.

2 WHATSAPP INTEL·LIGENT

En aquesta secció s'explica com està estructurada la implementació per portar a terme aquest projecte, com fun-

² Google Play és la plataforma per on es distribueixen les aplicacions en els dispositius Android.

cional i quines decisions han fet variar la planificació inicial.

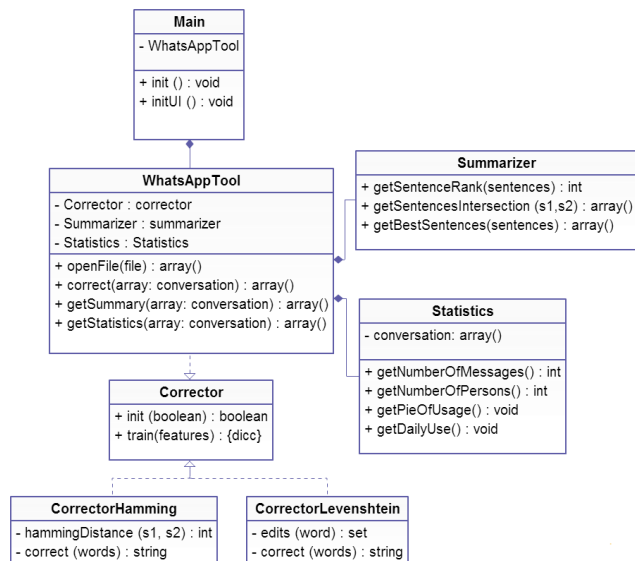
Aquest projecte està desenvolupat en el llenguatge Python versió 3.4, aquest llenguatge ofereix un gran ventall de possibilitats per realitzar un treball amb aquestes característiques, ja que compta amb grans llibreries científiques com Numpy, gràfiques com Matplotlib i relacionades amb la interpretació del llenguatge humà com NLTK (Natural Language Toolkit) i molta documentació al respecte.

El resultat final d'aquest treball és un client per a escriptori realitzat en Python (per tant compatible amb Windows, Linux i Mac) construït amb un paradigma de programació orientada a objectes i una interfície d'usuari per a fer més fàcil la seva utilització.

En les següents subseccions es mostra com està estructurada la part de la implementació del projecte i com funcionen els diferents mòduls que s'hi pot trobar.

2.1 Estructura de la implementació

La part implementada està construïda en diferents mòduls encarregats de diferents tasques que realitza el programa. Cada mòdul correspon a una classe per tal de fer el codi més fàcil de mantenir i entendre.



Il·lustració 2 Diagrama de classes de la implementació

El diagrama anterior mostra un esquema global del codi de l'aplicació, on es pot veure les diferents classes que formen el codi, amb els seus atributs i mètodes.

2.2 Mòdul corrector

Aquest mòdul és l'encarregat de comprovar la validesa de les paraules i corregir-les en els casos necessaris. Per tal de saber si una paraula és correcta en un determinat idioma, és necessari que "parlem" en aquell idioma. Aquest mòdul aprèn a parlar en català (idioma en què

s'han analitzat les converses) gràcies a un document que conté paraules i frases escrites correctament. Cal doncs llegir aquest fitxer i emmagatzemar-ne les paraules per una posterior anàlisi i comparació amb la conversa.

La qualitat d'aquest document té relació directa amb el resultat de la correcció. Per veure com afecta en el resultat final, a l'apartat de resultats es mostren un seguit de proves amb diferents mides de diccionari i com repercuteixen en el resultat final de la correcció.

Un cop el mòdul és capaç de reconèixer si una paraula està escrita correctament, ha d'intentar corregir aquelles que no trobi en el seu diccionari. Per tal de corregir aquesta paraula emprarem la mesura de la **distància**. Quan parlem de distància entre dues paraules, ens podem referir a la distància de **Hamming** [10] o a la distància de **Levenshtein** [11].

La distància de Hamming es defineix en informàtica com el nombre de caràcters o posicions diferents entre dues paraules o codis d'igual longitud. Per exemple la distància de Hamming entre les paraules "gat" i "gos" és igual a 2, ja que la segona i tercera lletra són diferents. Aquesta característica entre paraules es pot aplicar en les correccions si tenim en compte que el 90% dels casos, una paraula mal escrita i la seva correcció es troben a distància 1. Per tant, la classe encarregada de corregir paraules amb el mètode de Hamming intentarà trobar dins el diccionari una paraula amb la distància igual a 1.

La distància de Levenshtein es defineix com el nombre mínim d'edicions requerides per transformar una cadena de caràcters en una altra. Per exemple si tenim la cadena "paper" i la cadena "carrer" tenim una distància de Levenshtein de 3.

1. paper → caper (substitució de la 'p' per la 'c')
2. caper → carer (substitució de la 'p' per la 'r')
3. carer → carrer (inserció de la 'r' entre la 'r' i la 'e')

En aquest cas el corrector realitza transformacions a les paraules per tal d'aconseguir una paraula diferent i comprovar si aquesta està escrita correctament en el nostre diccionari.

2.3 Mòdul resum

El mòdul resum d'aquest projecte és l'encarregat de proporcionar un conjunt de frases perquè apareguin en el resum final de la conversa. Aquestes frases són les que es consideren més importants, per tant s'han d'escollir seguint un criteri i unes característiques.

El criteri que se segueix és la puntuació de les frases. Aquesta puntuació s'obté durant l'anàlisi de la conversa, on seguint unes característiques explicades a continuació, es dona un valor a cada frase. Les frases escollides són les que obtenen una puntuació més alta al final de l'execució.

La característica principal que se segueix és el nombre

de vegades que apareix una paraula en el text, és a dir la freqüència de cada paraula. Aquesta característica es coneix per tf-idf (term frequency-inverse document frequency) que ja s'ha vist en l'apartat d'estat de l'art i és utilitzat per cercadors, mineria de textos, etc. Aquesta puntuació que es dona a les paraules pel nombre de vegades que apareixen en el text és la que té més pes a l'hora d'escollir les frases finals. Aquesta decisió pot provocar alguns resultats inesperats en el resum, ja que es va comprovar que els termes que més apareixien en una conversa eren articles (el, la, els) o adverbis (sí, no) i aquests apareixien en el resum final sense proporcionar cap informació sobre la conversa. Per tal d'intentar puntuar més favorablement a altres paraules que potser no apareixien tantes vegades però tenien més significat que aquests articles, es va decidir elevar la longitud de la paraula avaluada al quadrat per tal de prioritzar les paraules més llargues.

Un altre punt a tenir en compte és la data del missatge, una intervenció feta al final de la conversa segurament té molta més importància que la primera proposta. La data del missatge ha d'afectar a la puntuació final d'aquesta. En algunes ocasions el primer missatge de la conversa és el que més informació aporta, ja que el creador difon entre els altres membres l'objectiu pel qual s'ha creat.

Finalment i després de l'obtenció de diversos resums, es va notar que hi havia paraules que apareixien molts cops en el resum, entre aquestes paraules hi podem trobar termes molt relacionats amb l'objectiu de la conversa. Per exemple si es tracta d'una conversa en què s'ha de decidir comprar un regal, la paraula "regal" dins de qualsevol frase, ha de tenir més pes. Aquest fet es pot definir com a **aprenentatge**, ja que s'ha obtingut a partir de diverses execucions i comparació de resultats.

2.4 Estadístiques

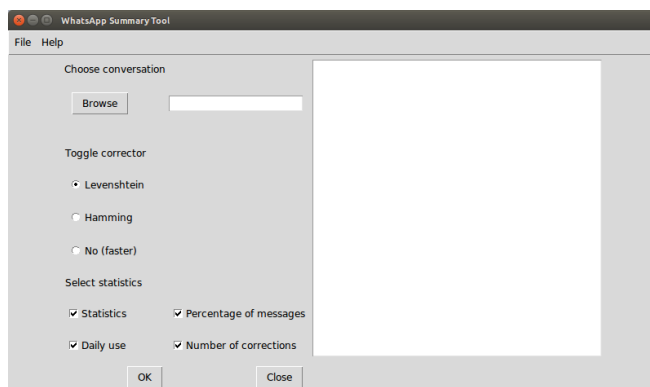
L'últim mòdul del client són les estadístiques. Aquesta part del programa s'encarrega, un cop finalitzats els altres mòduls, de recaptar tota la informació obtinguda fins al moment i mostrar-la de manera més ordenada i gràfica per tal que l'usuari pugui conèixer informació extra sobre la conversa.

Entre aquesta informació extra s'hi pot trobar informació sobre qui ha parlat més, el nombre de correccions, el nombre de correccions de cada persona, intervencions de cada persona en un moment determinat, etc.

2.5 Interfície d'usuari

Un dels requeriments del projecte és la usabilitat, per tant la pantalla amb la qual interactua l'usuari i el sistema ha de comptar amb un bon disseny. Aquesta interfície ha de permetre a l'usuari escollir la conversa, mostrar el resum final i escollir quina informació vol generar, per fer més personalitzable el resum.

En la il·lustració 3 es pot veure com l'usuari pot escollir la conversa amb un botó de navegació dins dels seus



Il·lustració 3 Captura de pantalla de la interfície en un sistema Linux

fitxers, pot escollir el mètode de correcció i finalment seleccionar aquelles estadístiques que vol generar. A la part dreta és on es mostrarà el resum de la conversa seleccionada.

3 EXPERIMENTS

En aquest apartat es mostren els diferents resultats obtinguts de la implementació del projecte. Es comença per mostrar els resultats obtinguts de l'execució del programa. En aquesta execució s'obtenen resums automàtics d'algunes converses i s'avalua la qualitat d'aquests.

Seguidament es realitzen diferents proves per fer una comparativa amb els temps d'execució, ja que és un requeriment d'aquesta implementació. Aquestes proves consisteixen a mirar com afecten els diferents paràmetres d'entrada del programa, com ara mida de la conversa o mida del diccionari i com afecten les diferents opcions que pot escollir l'usuari, com mètode de correcció i estadístiques.

Finalment es mostren les diferents estadístiques que es poden obtenir de l'execució.

Referent al tema dels resums, per tal de mostrar més àmpliament com funciona, s'ha destinat l'apartat d'apèndix a mostrar converses senceres i seguidament mostrar el resum obtingut. D'aquesta manera es pot comprendre més bé quin ha estat el fil principal de la conversa i avaluar nosaltres mateixos la qualitat del resum.

3.1 Resultats

El primer que s'avalua en l'apartat de resultats és l'obtenció del resum de la conversa. Com a part central del projecte cal controlar la qualitat d'aquest resum i no és una mesura fàcil de realitzar. L'objectiu final és que l'usuari pugui fer-se una idea de què ha tractat tota la conversa només amb la lectura de les intervencions més importants. Per tant, es pot avaluar el resum qualificant les intervencions que apareixen en el resultat final.

Per avaluar cada intervenció es pot puntuar de manera binària (ens serveix o no per entendre la conversa) o pon-

derar-la de l'1 al 10 segons la informació que es considera que ens ha aportat.

Per fer més significativa aquesta avaluació es va demanar a un conjunt d'usuaris que pertanyien a la mateixa conversa que ponderessin el resultat d'un resum i comentessin el resultat.

La conversa es va generar entre un grup d'usuaris per decidir què fer en un aniversari. La conversa sencera es troba a l'apartat A.1 de l'apèndix.

El resum és el següent:

7/5/2014 - Subjecte A 18:30: és dijous demà al vespre i nit.
6/5/2014 - Subjecte A 23:38: resum dijous aquest 8 maig sopar i festa, s'accepten propostes.
6/5/2014 - Subjecte B 23:28: jo no vindré, que divendres tinc una entrega i dijous no se a quina hora acabaré...
6/5/2015 - Subjecte C 22:40: si aquest dijous.
6/5/2015 - Subjecte A 22:34: eii, dijous 8 de maig

I l'opinió dels usuaris:

Subjecte A (nota 8): Apareix la paraula resum i a més a més concreta molt bé el dia en què es realitzarà l'esdeveniment. No apareix cap missatge irrelevant.

Subjecte B (nota 6): Només es mostren dies de la setmana, tot i que són útils es va parlar de moltes altres coses.

Subjecte C (nota 4): Amb aquest resum costa molt fer-se una idea dels temes que ha tractat la conversa.

Subjecte D (nota 7): És suficient per saber el dia en què es produeix el sopar i crec que és la dada més important que es pot extreure d'aquesta conversa.

Vista l'opinió dels usuaris es pot veure que un resum d'una conversa no significa el mateix per a totes les persones. Mentre per algunes és suficient en saber la data final que s'ha decidit, a d'altres els interessaria més que les intervencions del resum aportessin informació més diversa.

3.2 Mesures

Les primeres mesures que s'han realitzat són les més bàsiques de totes, l'anàlisi de diferents converses per obtenir-ne un resum. Durant la realització d'aquest treball, es va intentar optimitzar el temps d'execució, ja que en les primeres etapes el temps per a analitzar una conversa era inacceptable. Per tant, és important comprovar l'abast d'aquestes millores realitzant diverses proves amb diferents mides de conversa i comprovar com es comporta.

En la següent taula, s'exposen els temps d'execució total del client amb converses de diferent longitud. Aquestes converses són reals, però en les últimes proves es va controlar el nombre de missatges afegint diverses còpies

de converses amb una longitud menor. Per tant, la conversa de longitud 1500 missatges, és en realitat la suma de tres converses de 500 missatges. Aquest fet provoca una deformació del resum final, però no afecta el temps d'execució.

Taula 1 Temps d'execució del programa per a diferents longituds de conversa, mesurat en segons

Missatges	Levenshtein	Hamming	Sense corregir
10	0,2279	0,4129	0,0018
50	0,2422	1,0359	0,0209
100	0,3136	1,396	0,0541
300	0,6929	4,8115	0,4215
700	2,6066	9,8679	2,2497
1500	11,1942	31,5266	10,8709
3000	44,8075	75,5758	42,8732

En la taula anterior podem observar com es comporta el client per a diferents mides de conversa amb els diferents mètodes de correcció.

D'aquests resultats, es pot veure ràpidament que un dels mètodes de correcció, el de Hamming, requereix molt més temps, i que el mètode de Levenshtein gairebé no varia respecte a no corregir cap falta d'ortografia.

Aquests resultats poden sorprendre després d'haver llegit com funcionen els dos algorismes. Recordem que el mètode de Hamming només comprova les paraules d'una sola longitud i intenta trobar aquella amb la distància igual a 1, mentre que Levenshtein realitza diferents càlculs sobre la paraula per intentar trobar per força bruta³ una paraula escrita correctament.

Així doncs perquè existeix aquesta diferència tan notable?

Doncs a la pràctica, el que succeeix és que el corrector de Levenshtein realitza moltes menys comprovacions. La creació de moltes combinacions d'una paraula (esborrant, permutant o afegint) és molt més ràpid trobar una paraula que coincideixi amb el nostre diccionari. En canvi, el corrector per distància de Hamming, comprova totes les paraules d'una determinada longitud i només s'atura si en troba una amb distància igual a 1. Això té com a resultat que sigui molt més lent d'executar.

Per comprovar que realment el temps que diferencia les execucions és el de correcció, podem aprofundir una mica desglossant encara més el temps d'execució. En la següent taula es mostren els temps destinats a cada tasca (resumir i corregir) per als dos mètodes de correcció.

³ Un atac de força bruta consisteix en trobar una paraula provant totes les combinacions possibles.

Taula 2 Temps d'execució desglossat per tasques, mesurat en segons

1500 missatges	Temps resumir	Temps correcció	Total
Levenshtein	11,2958	0,2444	11,7785
Hamming	10,8108	16,1463	27,2191

En la taula anterior es pot comprovar que realment el que afecta més al temps final en el cas del Hamming, és el temps de correcció.

Amb aquests resultats es pot plantejar la pregunta, quin corrector és millor?

Doncs per tal d'intentar resoldre aquesta pregunta, a continuació llistarem un conjunt de paraules que contenen errors i quina solució ens proposa cada corrector.

Taula 3 Comparativa entre la correcció dels dos mètodes

Paraula correcta	Paraula escrita	Levenshtein	Hamming
adéu	adeu	ajeu	ageu
sopar	supar	sopar	sopar
després	despres	després	després
dijous	dijosu	dijous	-
perquè	perque	perquè	perquè
així	aixi	això	així
impossible	imposible	imposable	imposable
tothom	totom	tothom	tothom
sabeu	saveu	sabeu	haveu
genial	gebniat	genial	-
sorpresa	surpresa	sorpresa	sorpresa
següent	seguent	següent	segment
també	tambe	nambe	nambe
demà	dema	gema	rema
algú	algu	aliu	algú
això	aixo	això	així
llàstima	llastima	llàstima	llàstima
família	familia	familiar	família
pràctiques	practiques	pràctiques	pràctiques
merèixer	mereixer	merèixeré	merèixer
gintònic	gintonic	gintònic	gintònic
% encert		57,14	47,61

En la taula anterior es pot veure com actuen els dos correctors, en el resultat final veiem que el mètode de Levenshtein és més efectiu, això és degut al fet que l'altre és incapaç de corregir paraules en què per error s'hagi escrit una lletra de més o de menys, error molt comú en l'escriptura en dispositius mòbils.

Aquests correctors depenen d'un arxiu d'entrada que conté paraules i frases en la llengua en què volem corregir, en aquest cas en català. Aquest arxiu proporciona al corrector les paraules que considerem correctes, per tant, la qualitat d'aquest document és clau per tal de comprovar l'efectivitat en la correcció.

El temps que dedica el programa a la lectura d'aquest fitxer és molt petit (0,13 segons) per tant és aconsellable que aquest fitxer sigui el més complet possible.

Per tal de veure com afecta la mida del diccionari en el resultat final, s'han realitzat diverses proves modificant l'arxiu d'entrada per tal que contingui només el diccionari de català, diccionari més un llibre, o diccionari i dos llibres. El fet que contingui un diccionari proporciona la gran majoria de paraules i noms de la llengua. Els llibres ens proporcionen la gran majoria de formes verbals, articles, pronomes, etc. Les proves estan realitzades amb una conversa real d'aproximadament 75 missatges.

En la següent taula tenim el comportament del corrector de Levenshtein i de Hamming per a diferents mides de diccionari.

Taula 4 Comportament dels correctors per a diferents mides de diccionari

	Mida	Levenshtein		Hamming	
		Errors trobats	Correccions	Errors trobats	Correccions
Diccionari	1,15 MB	91	50	91	37
Dicc. + 1 llibre	1,53 MB	79	45	79	38
Dicc. + 2 llibres	1,80 MB	77	43	77	38

Podem veure que a mesura que incrementem la mida del diccionari, trobem menys errors, ja que paraules que estan escrites correctament no les coneixíem. Les correccions realitzades també disminueixen en relació als errors trobats.

En canvi el corrector de Hamming actua de manera diferent a l'anterior, realitza moltes menys correccions i a més a més, manté les correccions encara que la mesura del diccionari augmenti. Això passa perquè al corrector de Hamming li costa més trobar les correccions (ja que només corregeix quan la distància és 1) i encara que incrementem la mida del diccionari, no és capaç de trobar la correcció perquè els llibres afegits realment no aporten gaires paraules noves a un diccionari.

3.3 Estadístiques

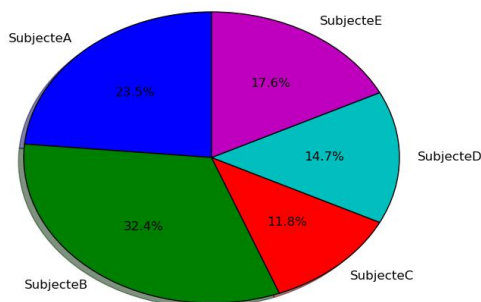
L'altra part important dels experiments són les estadístiques. A continuació es mostra el conjunt d'estadístiques que es poden obtenir a partir del client i com afecten en els resultats mostrats anteriorment.

Aquests gràfics serveixen per a obtenir més informació de la conversa i fer que sigui fàcil de veure amb un cop d'ull.

Aquestes imatges les podem trobar dins una carpeta creada en finalitzar l'execució del programa. L'usuari a partir de la interfície pot escollir quina informació vol generar.

3.3.1 Percentatge d'intervencions

A continuació veiem una de les estadístiques més senzilles, es tracta d'un diagrama de sectors que ens mostra en tant per cent de missatges que ha enviat cada usuari.



Il·lustració 4 Diagrama de sectors amb el percentatge de missatges de cada usuari.

Amb aquest diagrama podem saber qui ha parlat més en una conversa només visualitzant aquest gràfic.

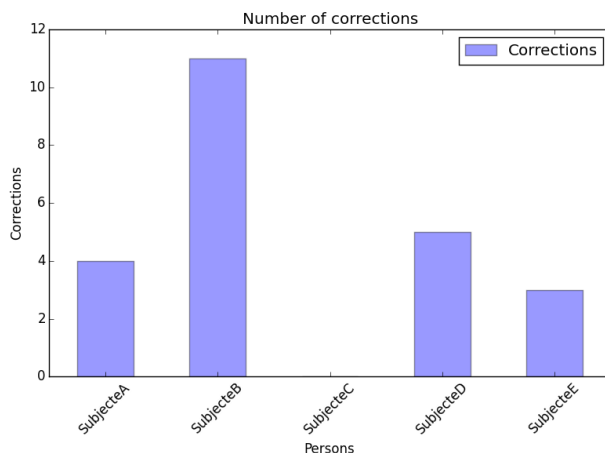
Aquesta opció també pot ser interessant per a altres tipus de conversa fora de l'abast d'aquest projecte. Per exemple es pot emprar aquesta característica per analitzar una conversa de milers de missatges amb un grup d'amics només per esbrinar qui ha parlat més en total⁴.

3.3.2 Correccions

Continuant amb les estadístiques més útils que es generen, trobem un diagrama de barres que ens mostra les correccions realitzades per a cada usuari.

La il·lustració 5 també proporciona una informació molt útil i ràpida de veure. Amb aquest esquema es pot veure quin usuari ha comès més faltes d'ortografia al llarg de la conversa.

⁴ Aquest anàlisi es podria fer sense corrector ja que per saber qui ha parlat més no es necessari a més a més que ens estalviariem el temps de correcció.

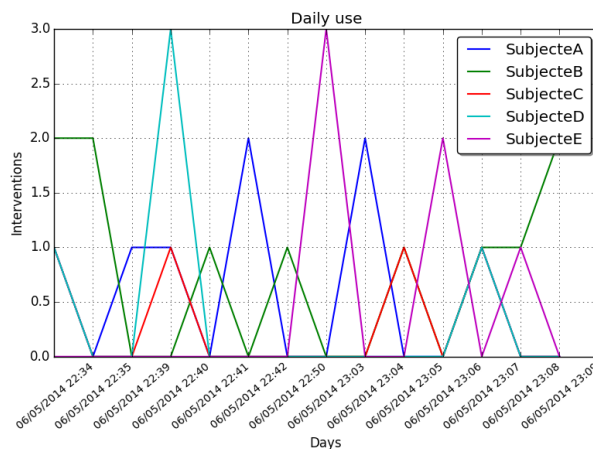


Il·lustració 5 Diagrama de barres en que es mostren les correccions que s'han fet per a cada usuari.

Similarment al diagrama de sectors anterior, aquesta característica també pot ser utilitzada per a converses més grans per tal d'obtenir informació de quin dels usuaris comet més errors.

3.3.3 Intervencions

El següent diagrama que es genera és una gràfica lineal que mostra la quantitat de missatges de cada persona en cada intervenció.



Il·lustració 6 Gràfic lineal en que es mostren les intervencions de cada usuari en diferents moments de la conversa.

Amb aquesta gràfica es pot percebre a quina hora s'ha parlat més i quins usuaris han intervingut més. Aquest gràfic ens permet veure quina ha estat l'hora en què s'han pres les decisions, ja que és normal que es produeixi un pic d'intervencions en què tothom estigui actiu i on es prenguin la majoria de les decisions. Si coneixem a quina hora s'ha parlat més, es pot revisar la conversa per centrar-se en aquella franja horària.

Aquesta estadística no ens proporciona una informació tan visual com les dues anteriors, ja que hem d'interpretar

una mica més la informació i pot ser una mica confusa.

3.3.4 Taula de dades

Finalment, es mostra una taula amb informació sobre la conversa. En aquesta taula es mostren dades que es poden extreure de l'anàlisi de la conversa.

Taula 5 Taula de dades resultant de les estadístiques

Nombre de persones	5
Número de missatges	34
Número de missatges processats	34
Nombre de correccions	23

En aquesta taula es mostra un resum de diferents aspectes de la conversa. Com que en aquest tipus de converses és normal trobar-hi elements que no són text, ens trobem que el sistema no ha processat alguns missatges o bé elements, com ara si un usuari entra en la conversa o en surt, tampoc no es processen però sí que compten com a intervenció.

3.3.5 Mesures en la generació d'estadístiques

La generació d'aquests gràfics es realitza un cop s'ha analitzat, corregit i proporcionat el resum de la conversa. Per tant aquesta generació també té un cost que no s'ha vist reflectit en les proves realitzades anteriorment.

Per veure com afecta en el temps total, s'han realitzat unes execucions i cronometrat el temps que trigaven a finalitzar.

En la següent taula es pot apreciar el temps d'execució per a diferents mides de conversa.

Taula 6 Temps d'execució de les diferents estadístiques mesurat en segons

Missatges	Taula	Sectors	Barres	Lineal	Temps total
10	0,3431	0,2295	0,2819	0,2954	1,644
300	0,3163	0,2362	0,2901	0,3770	2,1721
3000	0,3143	0,2381	0,2767	0,3834	46,6776

Com es pot apreciar, el temps per a generar aquestes imatges és menor d'1 segon i aquest temps no varia encara que incrementi la longitud de la conversa. Per tant és aconsellable generar sempre les estadístiques, ja que ens proporcionen més informació i gairebé no afecten el temps d'execució total (que és la suma del temps de totes les estadístiques, el temps de correcció i el temps per resumir).

4. CONCLUSIONS

El desenvolupament d'aquest projecte ha resultat satisfactori, s'ha aconseguit un producte funcional que ens permet obtenir un resum d'una conversa de WhatsApp i per tant, complir amb el principal objectiu del treball.

Durant el transcurs del projecte han sorgit canvis en els objectius proposats inicialment que han fet variar el producte final. La decisió d'aquests canvis es considera encertada vist el resultat final.

Finalment es poden enumerar les conclusions més importants que es poden extreure de tot el transcurs del treball:

La popularitat de WhatsApp: Des del moment d'escollir aquest Treball de final de grau fins ara, el fet de treballar amb una plataforma com és WhatsApp ha estat un tret motivador. Aquesta aplicació l'utilitza dia a dia un percentatge elevat de la població i la idea de realitzar un treball que pugui arribar a molta gent, sempre anima a millorar. Aquest concepte també es veu reforçat a l'hora d'explicar a familiars i amics en què consisteix aquest projecte, ja que la majoria coincideixen en dir que és molt interessant.

Qualitat de les mostres de converses: És sorprenent la quantitat de faltes d'ortografia que es poden trobar en les converses. Aquest fet no és només propi de l'aplicació WhatsApp, ni tan sols de la comunicació via dispositius mòbils que dificulta l'escriptura, sinó que generalment la qualitat de les comunicacions informals és força baixa. La comunicació per SMS⁵ va suposar un canvi en la manera de comunicar-nos, ja que aquests tenien una longitud limitada i afavorien l'ús d'un llenguatge reduït. Aquest llenguatge encara s'utilitza molt per a escriure de manera ràpida i còmode. En contrast a tot això, actualment també és l'època en què disposem de més informació i de més eines per tal d'escriure correctament.

La complexitat del corrector: La dificultat des de bon principi amb la codificació dels caràcters de la llengua catalana en el llenguatge de programació Python, ja va significar un contratemps des del començament. Seguidament, ja es va preveure que la correcció de les paraules mal escrites comportaria una gran part del desenvolupament. Després de dedicar moltes hores a aquest mòdul, encara ara és una eina molt rudimentària i el resultat no sempre és el desitjat.

Algorisme principal: És complicat decidir com realitzar el resum. En diverses ocasions s'ha canviat la fórmula que puntua les frases perquè tenia un efecte molt diferent depenent de la conversa. Finalment el resultat obtingut no aporta el contingut suficient per a entendre completament el resultat de la conversa, això és degut al fet que la gran majoria de converses s'escapen de les condicions òptimes per les quals està pensat aquest projecte.

Caràcter informal de les converses: Aquest projecte ha estat desenvolupat agafant converses reals de WhatsApp. La gran majoria de converses que resideixen a la nostra aplicació són informals. Aquest fet condiciona el tipus de text que trobarem al seu interior. Tot i ser una conversa

⁵ SMS és l'acrònim per a servei de missatges curts (Short Message Service)

curta i amb un objectiu molt clar, s'hi pot trobar tot tipus de material irrellevant que dificulta la tasca del resum. Per aquesta raó, s'ha optat per crear diverses converses fictícies on només hi ha informació útil.

El llenguatge Python com a eina de desenvolupament: L'elecció d'aquest llenguatge no va ser trivial, en gran part per l'experiència prèvia a aquest treball però també per la gran quantitat de documentació relacionada amb projectes d'aquest tipus. En les primeres etapes es va fer recerca d'altres llenguatges i no es va trobar la mateixa informació. En etapes més avançades, la decisió de canviar els objectius, la creació del client amb la llibreria "Tk" de Python i la generació de gràfics amb la llibreria "matplotlib" han fet encara més remarcable que va ser una bona decisió.

Qualificació dels resultats obtinguts: Gràcies al fet que un punt important en el desenvolupament ha estat el rendiment, el temps d'execució es considera acceptable vist la longitud de les converses que es pretenen tractar en aquest projecte. També pot ser utilitzat en grans converses de milers d'intervencions, simplement per generar les estadístiques i saber qui ha parlat més en converses amb anys d'antiguitat.

4.1 POSSIBLES AMPLIACIONS I MILLORES

En aquest apartat s'anomenaran algunes de les línies que es poden seguir en aquest projecte després del desenvolupament fet per aquest treball de fi de grau.

- **Aplicació per a dispositius mòbils:** Tot i l'aparició de WhatsApp per la Web, la gran majoria del temps utilitzem l'aplicació des del mòbil. Per tant aquest servei que es proporciona tindria molt de sentit si es pogués fer des del mateix dispositiu.
- **Millores en el corrector i en el sistema de resums:** Com s'ha pogut apreciar en l'apartat de resultats, els mòduls de correcció i resums són eines força rudimentàries si les comparem amb les altres que trobem en el mercat i utilitzem cada dia. Algunes d'aquestes eines són proporcionades per grans empreses amb un gran equip de treballadors. Per tant sempre es pot millorar la qualitat d'aquestes parts per tal de proporcionar millor resultat.
- **Proporcionar més estadístiques:** L'apartat d'estadístiques sempre es pot ampliar. Actualment es mostren dades que poden ser d'interès per a tots els usuaris, però també és possible que alguns usuaris necessitin obtenir altre tipus d'informació més específica i així la part de la interfície d'usuari mostri més opcions per escollir i sigui més personalitzable.

AGRAÏMENTS

Agrair a Joan Serra Sagristà per l'ajuda proporcionada com a director del treball.

Als usuaris que han proporcionat converses per tal de tenir més material amb el qual treballar.

Als usuaris que han dedicat temps per utilitzar el programa i aportar el seu punt de vista com a usuaris finals.

BIBLIOGRAFIA

- [1] WhatsApp Blog, *WhatsApp Web*. [consultat: 08 Febrer 2015]. Disponible a Internet: <https://blog.whatsapp.com/614/WhatsApp-Web>
- [2] Android Appetizers, *WhatStat*. [Aplicació Android]. Disponible a Internet: <https://play.google.com/store/apps/details?id=com.androidappetizers.whatstat>
- [3] David M. Zajic, Bonnie J. Dorr, Jimmy Lin. 2008. *Single-Document and Multi-Document Summarization Techniques for Email Threads Using Sentence Compression*. Department of Computer Science, College of Information Studies, University of Maryland, College Park.
- [4] Gabriel Murray, Giuseppe Carenini. 2008. *Summarizing Spoken and Written Conversations*. Department of Computer Science, University of British Columbia, Vancouver, Canada.
- [5] Guanting Tang, Jian Pei, Wo-Shun Luk. 2013. *Email Mining: Tasks, Common Techniques, and Tools*. School of Computing Science, Simon Fraser University, Burnaby BC, Canada.
- [6] Michael Dittenbach. 2010. *Scoring and Ranking Techniques - tf-idf term weighting and cosine similarity*. [consultat: 25 Gener 2015]. Disponible a Internet: <http://www.ir-facility.org/scoring-and-ranking-techniques-tf-idf-term-weighting-and-cosine-similarity/>
- [7] Giuseppe Carenini, Raymond T. Ng, and Xiaodong Zho. 2007. *Summarizing email conversations with clue words.*, pàgines de 91 a 100. Alberta, Canada.
- [8] Owen Rambow, Lokesh Shrestha, John Chen, and Chirsty Lauridsen. 2004. *Summarizing Email Threads*. Pàgines de 105 a 108, Stroudsburg, Pennsilvània, EUA. Associació de lingüística Computacional.
- [9] Shlomi Babluki, *Build your own summary tool!* [consultat: 17 Gener 2015]. Disponible a Internet: <http://thetokenizer.com/2013/04/28/build-your-own-summary-tool/>
- [10] P.Danziger, *Hamming distance* [consultat: 5 Febrer 2015]. Disponible a Internet: <http://www.math.ryerson.ca/~danziger/professor/MTH108/Handouts/codes.pdf>
- [11] Stanford University, *Minimum Edit Distance Definition of minimum edit distance* [consultat: 8 Febrer 2015]. Disponible a Internet: <https://web.stanford.edu/class/cs124/lec/med.pdf>

APÈNDIX

En l'apartat d'apèndix es mostra el resum resultant per a diferents tipus de converses. Les converses són completes per tal de poder mostrar tots els missatges que contenen i seguidament es mostrarà el resum. Amb això és pretén intentar mostrar com funciona el programa. També es mostra com actua el corrector o com aquest fa variar el resultat del resum. Les converses no han sigut corregides.

A.1 CONVERSA REAL 1

6 May 22:34 - Subjecte A: Eii, dijous 8 de maig
 6 May 22:34 - Subjecte A: Festa precumple Subjecte H (9 cumple Subjecte H)
 6 May 22:35 - Subjecte A: Sopar + festa
 6 May 22:35 - Subjecte A: S'accepten propostes x on nar
 6 May 22:37 - Subjecte A changed this group's icon
 6 May 22:39 - Subjecte B: jajajaj
 6 May 22:40 - Subjecte B: es aqest
 6 May 22:40 - Subjecte C: Quina por hem feu... XD
 6 May 22:40 - Subjecte H: Si aquest dijous
 6 May 22:40 - Subjecte H: Aixi fare els 23 de festa 😊
 6 May 22:41 - Subjecte B: jajajajha
 6 May 22:42 - Subjecte B: uaaa anem a bikiiniiii
 6 May 22:50 - Subjecte A: • 🍷 • • •
 6 May 23:03 - Subjecte G: • 🍷 ☐
 6 May 23:04 - Subjecte A: Bikini gratis??
 6 May 23:05 - Subjecte G: Suposu ke si hi ha llista si
 6 May 23:10 - Subjecte H: Difuneu el missatge!
 6 May 23:12 - Subjecte A: Ss, digueu i agreguem al ki sigui més a aquest gruuup!!!
 6 May 23:22 - Subjecte H: Ara els afegeix el Subjecte A
 6 May 23:24 - Subjecte G: Putse
 6 May 23:26 - Subjecte D: Afegiu al Subjecte J
 6 May 23:26 - Subjecte D: Tot i que diu que té classe divendres al matí, però potser s'anima
 6 May 23:27 - Subjecte G: Cert
 6 May 23:27 - Subjecte G: Benu es pot fe tot
 6 May 23:28 - Subjecte D: Jo no vindré, que divendres tinc una entrega i dijous no se a quina hora acabaré...
 6 May 23:28 - Subjecte D: Però ja pensaré a felicitar-te Subjecte H!! Hihhi!!
 6 May 23:28 - Subjecte D: Passeu-ho molt bé!!
 6 May 23:29 - Subjecte I was added
 6 May 23:30 - Subjecte J was added
 6 May 23:31 - Subjecte K was added
 6 May 23:38 - Subjecte A: Resum: dijous aquest (8maig) sopar... + festa->bikini, s'accepten propostes
 7 May 00:20 - Subjecte G: A mi tan mes po plegu a les 9
 7 May 00:20 - Subjecte G: ☐
 7 May 00:23 - Subjecte H: ☐ ☐
 7 May 07:17 - Subjecte I: Jo no
 7 May 08:45 - Subjecte D left
 7 May 09:20 - Subjecte K: Jo no puc, mhe tornat tu Subjecte A i he danar a classe lendema..... ☐
 7 May 10:12 - Subjecte A: Jajaja, jo tinc classe l'endema XD i hi aniré
 7 May 10:33 - Subjecte F: (a les 2)
 7 May 10:34 - Subjecte A: ☐

7 May 10:35 - Subjecte F: Jo ja us vaig dir, vinc a sopar i a les 12 cap a casa
 7 May 10:35 - Subjecte E: Jo vindré una estoneta, que l'endema tinc exercici avaluable
 7 May 16:56 - Subjecte C: Jo no puc, divendres he de pintar l'habitació per poder tornar a pinar el dissabte i que dilluns estigui ben sec
 7 May 17:02 - Subjecte G: ☐
 7 May 17:09 - Subjecte B: ostiii ami tampoc em va gairee beee dv
 7 May 17:09 - Subjecte B: avui ja he empelmat
 7 May 17:09 - Subjecte B: jajaa
 7 May 17:21 - Subjecte G: Po no es dijous?
 7 May 17:22 - Subjecte G: Si es div jo tinc entrenu
 7 May 17:24 - Subjecte G: Subjecte B tu de sense dormir
 7 May 18:03 - Subjecte A: És dijous (demà) al vespre i nit
 7 May 18:18 - Subjecte B: jajajaja ami no em ca gairee beee
 8 May 15:11 - Subjecte E: Quin és el planning al final??
 8 May 15:20 - Subjecte F: Això això
 8 May 15:51 - Subjecte C: Jo tinc feina, hem sap greu
 8 May 18:18 - Subjecte J: Jo tampoc puc

A.1.1 RESUM

7/5/2014 - Subjecte A 18:30: és dijous demà al vespre i nit.
 6/5/2014 - Subjecte A 23:38: resum dijous aquest 8 maig sopar i festa, s'accepten propostes.
 6/5/2014 - Subjecte B 23:28: jo no vindré, que divendres tinc una entrega i dijous no se a quina hora acabaré...
 6/5/2015 - Subjecte C 22:40: si aquest dijous.
 6/5/2015 - Subjecte A 22:34: eii, dijous 8 de maig

A.2 CONVERSA REAL 2

30 Jan 13:43 - Subjecte A: Dema desteueixo el grup
 30 Jan 13:43 - Subjecte A: No us estresseu xD
 30 Jan 13:43 - Subjecte B: XDDDD millor
 30 Jan 13:43 - Subjecte A: Diu la meva mama que podem fer els pastisos i tal avui, però el Subjecte X està d'entre-gues
 30 Jan 13:44 - Subjecte A: Creieu que el podem arrancar de casa?
 30 Jan 13:44 - Subjecte C: I te classes avui i dema
 30 Jan 13:44 - Subjecte A: Hahahahha
 30 Jan 13:44 - Subjecte B: Voleu venir demà després de sopar a casa a fer postees?
 30 Jan 13:44 - Subjecte A: ☺
 30 Jan 13:44 - Subjecte B: A sopar no puc venir demà, ni avui tampoc
 30 Jan 13:44 - Subjecte A: Saps de quina hora a quina hora?
 30 Jan 13:45 - Subjecte C: N pi si n avaba les entregues aura d suspendre les classes
 30 Jan 13:45 - Subjecte B: Demà a la tarda ha dit que de 4 a 8:30
 30 Jan 13:45 - Subjecte A: Tampoc dormira xD
 30 Jan 13:45 - Subjecte A: Dic avui hahaha
 30 Jan 13:46 - Subjecte A: Tambe tenia una altra idea, fem el pastis, i quan estigui fet, l'anem a pesca allà on sigui i li

portem

30 Jan 13:46 - Subjecte B: Potser és complicat organitzar-li algo sorpresa tal com té el cap de setmana
 30 Jan 13:46 - Subjecte A: En plan surprise🤔
 30 Jan 13:46 - Subjecte A: Si
 30 Jan 13:47 - Subjecte B: I si li diem Subjecte X quan vols celebrarho i que ho decideixi ell?
 30 Jan 13:50 - Subjecte A: Com vulgeu... A mi em ve de gust fer-li algo que no s'espera... Això ultim que he dit no cal robar-li mes de 5 minuts, i despres celebrar-ho com vulgui ell
 30 Jan 13:50 - Subjecte B: Xi, això és guai
 30 Jan 13:50 - Subjecte B: I si li dic al meu germà que li demani una hora... I li fem una sorpresa?
 30 Jan 13:51 - Subjecte B: Però potser li diu que sí a ell i que no a un altre i despres hi perd diners...
 30 Jan 13:51 - Subjecte A: Pode s'estressa més si li demana mes feina... Jo ho havia pensat així:
 30 Jan 13:51 - Subjecte A: Quan estigui el pastis fet, el truquem: Subjecte X on ets?
 30 Jan 13:52 - Subjecte A: A) a casa, fent entregues. Anem 5 min a casa seva
 30 Jan 13:52 - Subjecte A: B) anant a fer classes. L'anem a pescar pel carrer
 30 Jan 13:52 - Subjecte A: C) a classe. L'esperem a la sortida
 30 Jan 13:52 - Subjecte A: Ca us sembla?☐
 30 Jan 13:53 - Subjecte B: O igualment si s'ha de refredar el pastís podem anarhi més tard
 30 Jan 13:53 - Subjecte B: A mi bé
 30 Jan 13:53 - Subjecte B: Li podem escriure Subjecte X amb xocolata al pastís
 30 Jan 13:53 - Subjecte A: Hehehehebe si
 30 Jan 13:53 - Subjecte A: Aviam que hi diuen els altres
 30 Jan 13:54 - Subjecte B: A mi m'està tot bé 😊
 30 Jan 13:58 - Subjecte C: Jo n e d fer res per tant m'està b
 30 Jan 13:59 - Subjecte A: 🙌🙌
 30 Jan 14:20 - Subjecte D: Guai! Voto sí
 30 Jan 14:26 - Subjecte A: Bieeen
 30 Jan 14:28 - Subjecte A: Ara vaig a comprar amb la meva mama
 30 Jan 14:28 - Subjecte A: Quin pastis us ve de gust
 30 Jan 14:29 - Subjecte C: El k vulgis i sigi facil
 30 Jan 14:30 - Subjecte B: A mi tb m'és igual!
 30 Jan 14:31 - Subjecte A: La Laura de Huelva, qur sabia fer un pastis molt bo, però no em respon per la recepta
 30 Jan 14:32 - Subjecte A: Brownie? Pastis de poma?
 30 Jan 14:32 - Subjecte B: Aaaaah
 30 Jan 14:32 - Subjecte B: Pastís de mandarina amb xocolata blanca?
 30 Jan 14:33 - Subjecte A: Becs xD
 30 Jan 14:34 - Subjecte B: És recepta de la meu mare... És molt bo...
 30 Jan 14:39 - Subjecte B: 😊😊😊
 30 Jan 14:39 - Subjecte C: Xdd
 30 Jan 14:39 - Subjecte D: Brownie!
 30 Jan 14:40 - Subjecte A: tu, xocolata amb fruita?
 30 Jan 14:40 - Subjecte D: Nah, em veritat qualsevol em semblarà bo jajaja
 30 Jan 14:40 - Subjecte A: Es qur no m'agrada la xocolata

blanca

30 Jan 14:40 - Subjecte A: El de poma, es com les pastes de poma del forn
 30 Jan 14:40 - Subjecte A: Facil i barat
 30 Jan 14:40 - Subjecte A: Pode em decanto per aquest jo
 30 Jan 14:41 - Subjecte B: A mi tb qualsevol
 30 Jan 14:41 - Subjecte B: De fet és pel Subjecte X no? Què li agrada a ell?
 30 Jan 14:41 - Subjecte A: Doncs que sigui brownie🍪🙌
 30 Jan 14:42 - Subjecte A: Mmmh normalment no us ho demanaria, pero vaig curta de pasta... Us sbla be si pagem els ingredients entre tots? Se que es idea meva eh
 30 Jan 14:42 - Subjecte A: I si no us va be, me'n faig carreg
 30 Jan 14:42 - Subjecte A: 😊
 30 Jan 14:43 - Subjecte B: Per mi prou, si ho paguem entre tots ens sortirà a menys dun euro x cap
 30 Jan 14:43 - Subjecte B: Tinc una recepta de brownie molt bona si voleu
 30 Jan 14:43 - Subjecte B: Me lhan demanada gent i tot despres de tastarlo 😊
 30 Jan 14:48 - Subjecte A: Doncs digues els ingredients
 30 Jan 14:48 - Subjecte A: 🍪🍪🙌
 30 Jan 14:51 - Subjecte A: ❤️
 30 Jan 14:51 - Subjecte B: Vainilla en tinc a casa
 30 Jan 14:51 - Subjecte B: X si no en voleu comprar x tan poca
 30 Jan 14:53 - Subjecte B: I nous rn tinc del jardí... Però sense trencar 😊
 30 Jan 14:54 - Subjecte D: Anemhi tots a comprar, si voleu
 30 Jan 14:55 - Subjecte A: Ah! Tambe
 30 Jan 14:56 - Subjecte A: Jo soc ara al mercadona
 30 Jan 14:56 - Subjecte B: Home, si ja hi ets...
 30 Jan 14:56 - Subjecte B: No cal que fem tants volts no? Meno, si et fa anar malament hi tornem després
 30 Jan 14:57 - Subjecte B: Però sino... ☐
 30 Jan 14:57 - Subjecte A: Jo tenia pensat comprar-ho ara
 30 Jan 14:57 - Subjecte A: Pero si hi voleu anar despres, fem
 30 Jan 14:57 - Subjecte B: Jo si a tu no et fa anar malament m'estimo més ara
 30 Jan 14:58 - Subjecte B: Així no anirem tant de cul a la tarda i no hem d'anar a fer volts amb el cotxe
 30 Jan 14:58 - Subjecte A: Ok
 30 Jan 14:58 - Subjecte B: Però sino després
 30 Jan 14:58 - Subjecte B: Les nous i la vainilla ja ho poso jo si voleu
 30 Jan 14:59 - Subjecte A: Uk
 30 Jan 15:24 - Subjecte A: He agafat mantega i xocolata fondant
 30 Jan 15:24 - Subjecte A: Farina i sucre en tinc a casa
 30 Jan 15:25 - Subjecte A: Falta res?
 30 Jan 15:26 - Subjecte B: Canyella
 30 Jan 15:26 - Subjecte B: Crec que en tinc
 30 Jan 15:26 - Subjecte A: Ah si, jo tm
 30 Jan 15:26 - Subjecte B: Pos ja està
 30 Jan 15:27 - Subjecte B: El motllo del cor és a Cerdanyola?
 30 Jan 15:32 - Subjecte A: Si, pero crec q la meva mama en va comprar un altrw
 30 Jan 15:33 - Subjecte B: 😊😊😊

30 Jan 15:33 - Subjecte B: A casa en tinc d'aquells de sili-
cona
30 Jan 15:33 - Subjecte B: Al final nem a casa teva o a la
meva?
30 Jan 15:36 - Subjecte A: Meva si voleu
30 Jan 15:36 - Subjecte A: La meva mama diu q cap pro-
blem
30 Jan 15:36 - Subjecte B: A mi m'està bé
30 Jan 15:37 - Subjecte B: A casa meva no hi haurà ningú
per això, si enredem a la teva mare
30 Jan 15:37 - Subjecte A: No pas
30 Jan 15:57 - Subjecte D: 😊
30 Jan 15:57 - Subjecte B: Ooooooh
30 Jan 15:58 - Subjecte A: Son 2 euros, ho agafo?
30 Jan 15:59 - Subjecte B: Mmmmmmm
30 Jan 15:59 - Subjecte D: Com us sembli
30 Jan 15:59 - Subjecte B: Amb brownie no sé si s'hi en-
ganxarà...
30 Jan 16:26 - Subjecte D: A quina hora voleu quedar 🍷?
30 Jan 16:27 - Subjecte A: A partir de les 5 podeu venir
30 Jan 17:00 - Subjecte B: Jo només he d'agafar el cotxe
30 Jan 17:01 - Subjecte D: Jo vestir-me
30 Jan 17:02 - Subjecte C: Jo ja soc aki
30 Jan 17:02 - Subjecte B: Ous en tens Subjecte A?
30 Jan 17:07 - Subjecte B: Subjecte C teniu ous pel pastís?
30 Jan 17:08 - Subjecte D: Jo en tinc sino
30 Jan 17:08 - Subjecte D: Diria jajaja
30 Jan 17:09 - Subjecte B: Jo tb
30 Jan 17:16 - Subjecte B: Ja estic a punt però no sé si aga-
far els oooooos! 😊😊😊
30 Jan 17:16 - Subjecte D: Jajajajaja
30 Jan 17:18 - Subjecte B: Subjecte D tens pintaungles
blau?
30 Jan 17:19 - Subjecte C: XD
30 Jan 17:19 - Subjecte C: Si, en tinc 3👉
30 Jan 17:19 - Subjecte B: Subjecte C ous.sí ous.no?
30 Jan 17:19 - Subjecte B: Però en porto o no?
30 Jan 17:19 - Subjecte C: Ai xD soc el subjecte A
30 Jan 17:20 - Subjecte C: Nooo
30 Jan 17:20 - Subjecte B: Hahaha vale
30 Jan 17:24 - Subjecte D: Si qie en tinc!
30 Jan 17:25 - Subjecte B: De pintaungles o d'ous?
30 Jan 17:25 - Subjecte D: Pintaungles
30 Jan 17:25 - Subjecte D: I ous tambe en tenim jaja
30 Jan 17:27 - Subjecte B: 😊😊😊😊
30 Jan 17:27 - Subjecte B: Ja estic, quan vulguis passo a
buscarte
30 Jan 17:30 - Subjecte D: Jo tambe estic!
30 Jan 17:32 - Subjecte B: Pos ja pots baixar!
30 Jan 17:34 - Subjecte D: Oqq!
31 Jan 10:42 - Subjecte A: Compren el perepunyetes?
31 Jan 10:59 - Subjecte B: Siiiiiii
31 Jan 10:59 - Subjecte B: Voleu que li demani al meu pare
si ens el ven a preu de cost?
31 Jan 11:05 - Subjecte A: Xiii😊☐☐
31 Jan 11:59 - Subjecte E: quins son els plans davui doncs?
31 Jan 12:39 - Subjecte C: Anar a sopar suposo pk el Sub-
jecte x fins les 8 i 30 estara ffent classes
31 Jan 12:40 - Subjecte C: Tt i k si voleu kdar per fer algo
jo per alla a les 5 kuarts de 6 estare lliure

31 Jan 12:42 - Subjecte E: vale 🍷
31 Jan 12:42 - Subjecte E: jo sopar i prou
31 Jan 14:11 - Subjecte B: On sopem per cert? S'ha pronun-
ciat monsieur X?
31 Jan 14:11 - Subjecte A: Amb mi nope
31 Jan 14:12 - Subjecte D: No sabem res
31 Jan 14:13 - Subjecte B: Hmmmm
31 Jan 21:16 - Subjecte B: Tinc el Perepunyetes 🍷
31 Jan 21:23 - Subjecte C: Guaii!! 😊😊
31 Jan 21:23 - Subjecte C: ☐
31 Jan 21:23 - Subjecte A: Quan surt per cap?
31 Jan 21:23 - Subjecte B: Mmmm
31 Jan 21:23 - Subjecte B: 1
31 Jan 21:23 - Subjecte C: Jeje ok
31 Jan 21:24 - Subjecte B: Som 5 a repartir oi?
31 Jan 21:24 - Subjecte C: Sui
31 Jan 21:24 - Subjecte C: Si
31 Jan 21:24 - Subjecte B: 😊

A.2.1 Resum

30/1/2015 17:19 - Subjecte C: ai xd soc el subjecte a
30/1/2015 17:18 - Subjecte B: subjecte d tens pintaungles
blau?
30/1/2015 17:02 - Subjecte B: ous en tens subjecte a?
30/1/2015 13:53 - Subjecte B: li podem escriure subjecte x
amb xocolata al pastís
30/1/2015 13:51 - Subjecte A: quan estigui el pastis fet, el
truquem subjecte x on ets?

A.3 Conversa 3

22 Sep 20:20 - Subjecte A created group "Sopar Aniversa-
ri"
22 Sep 20:28 - Subjecte A added you
22 Sep 20:21 - Subjecte A: Barcelonins! Si voleu, podeu
anar fins a la UAB i us passem a buscar amb cotxe allà
22 Sep 20:21 - Subjecte B: M has trucat?
22 Sep 20:21 - Subjecte B: El Subjecte C i jo anavem a pillar
el fgc
22 Sep 20:21 - Subjecte B: Fins a santcu
22 Sep 20:23 - Subjecte C: Jo igual
22 Sep 20:24 - Subjecte A: He parlat b el Subjecte C
22 Sep 20:24 - Subjecte A: Li he proposat que vingueu fins
a la uab i us passem a buscar amb cotxe
22 Sep 20:25 - Subjecte C: Bueno com vulguis pero es anar
mes lluny, no?
22 Sep 20:25 - Subjecte D: Una mica xd
22 Sep 20:25 - Subjecte A: Subjecte B, Subjecte D i el Sub-
jecte C seran a les 20:40 a cat, podeu quedar i anahi junts
22 Sep 20:25 - Subjecte D: Esperens Subjecte C!
22 Sep 20:25 - Subjecte A: Si, pero aixi hi anem amb cotxe i
ale
22 Sep 20:26 - Subjecte C: Com q a cat?
22 Sep 20:26 - Subjecte B: I
22 Sep 20:26 - Subjecte B: Pl cat
22 Sep 20:27 - Subjecte C: Jo he d'anar fins a gracia, no cal
que baixi fins cat per despres pujar, no? XD
22 Sep 20:27 - Subjecte C: Podem quedar a gracia
22 Sep 20:27 - Subjecte B: Doncs t avisem quan passi eq
ferro o algo?
22 Sep 20:27 - Subjecte C: Ci

22 Sep 20:27 - Subjecte C: Jo vaig tirant
 22 Sep 20:27 - Subjecte D: Jajaja que las anat a pillar a gracia?
 22 Sep 20:28 - Subjecte C: Home xD
 22 Sep 20:28 - Subjecte B: Home Subjecte C si vls qdar amb ell jo he d anar a cat per nassos
 22 Sep 20:30 - Subjecte D: Jo soc a passeig ja xd
 22 Sep 20:32 - Subjecte D: • • • 🐶🐶🐶
 22 Sep 20:34 - Subjecte C: No em surt a compte baixar xD
 22 Sep 20:35 - Subjecte B: Jo a pl cat
 22 Sep 20:35 - Subjecte C: Jo arribo en 5 minuts a gracia
 22 Sep 20:36 - Subjecte C: Llavors anem fins la Uab?
 22 Sep 20:37 - Subjecte D: Sí uab
 22 Sep 20:37 - Subjecte C: Oks
 22 Sep 20:42 - Subjecte C: Jo ja hi doc
 22 Sep 20:42 - Subjecte C: Soc
 22 Sep 20:42 - Subjecte C: Ja direu xD
 22 Sep 20:43 - Subjecte A: Ja ets a la uab??
 22 Sep 20:43 - Subjecte C: Nooo
 22 Sep 20:43 - Subjecte C: A gracia
 22 Sep 20:43 - Subjecte C: XD
 22 Sep 20:44 - Subjecte B: Esperantnos?
 22 Sep 20:45 - Subjecte C: Sii
 22 Sep 20:45 - Subjecte D: Subjecte C serem al ultim vagó
 22 Sep 20:45 - Subjecte C: No tio al primer
 22 Sep 20:45 - Subjecte C: XD
 22 Sep 20:46 - Subjecte B: Ultim
 22 Sep 20:46 - Subjecte C: Bah xD
 22 Sep 20:46 - Subjecte C: Guaeume lloc
 22 Sep 20:46 - Subjecte B: Ja esta xd
 22 Sep 20:47 - Subjecte A: Aaah
 22 Sep 20:47 - Subjecte A: Jajaja
 22 Sep 20:47 - Subjecte C: A la uab es surt pel primer... ☐
 22 Sep 20:47 - Subjecte A: ☐
 22 Sep 20:48 - Subjecte C: S1?
 22 Sep 20:48 - Subjecte D: ☐
 22 Sep 20:48 - Subjecte C: No? XD
 22 Sep 20:50 - Subjecte C: XD
 22 Sep 20:53 - Subjecte D: Ara arrivem a gracia
 22 Sep 20:53 - Subjecte D: Amb el de sabadell
 22 Sep 20:53 - Subjecte C: Okeis
 22 Sep 20:54 - Subjecte D: Recorda ultim vagó ☐
 22 Sep 20:54 - Subjecte C: Pillo el següent doncs
 22 Sep 21:07 - Subjecte A: Per on neu?
 22 Sep 21:09 - Subjecte C: Floresta crdc
 22 Sep 21:09 - Subjecte C: Crec
 22 Sep 21:14 - Subjecte C: San cugat
 22 Sep 21:17 - Subjecte A: Quedem a sota de plaça civica, on et deixem amb el Subjecte D a vegades,
 22 Sep 21:17 - Subjecte C: Okiiis
 22 Sep 21:25 - Subjecte C: Ja hi som
 22 Sep 22:32 - Subjecte E: Holaaa ja estiic
 22 Sep 22:32 - Subjecte E: Què feu?
 22 Sep 22:32 - Subjecte F: Estem aqui que ara em acavat
 22 Sep 22:33 - Subjecte F: I farem postres
 22 Sep 22:33 - Subjecte E: Poooos us espero a casa
 22 Sep 22:33 - Subjecte E: I aixi sopo
 22 Sep 23:00 - Subjecte A: Vale
 22 Sep 23:01 - Subjecte C: Molt malament
 22 Sep 23:03 - Subjecte E: Subjecte C em convides a una

birra eh

22 Sep 23:03 - Subjecte D: Si k et convido

A.3.1 Resum

22/9/2015 20:45 - Subjecte D: subjecte c serem al ultim vagó

22/9/2015 20:28 - Subjecte B: home subjecte c si vls qdar amb ell jo he d anar a cat per nassos

22/9/2015 20:25 - Subjecte D: esperens subjecte c!

22/9/2015 20:24 - Subjecte A: he parlat b el subjecte c

22/9/2015 20:21 - Subjecte B: el subjecte c i jo anavem a pillar el fgc